PROBABILIDADES Y ESTADÍSTICA (C)
Práctica 0
Introducción a R

1. a) Sea

   ```
   x <- c(1,2,3)
   y <- c(6,5,4).
   ```

   Predecir el resultado de correr cada una de las siguientes líneas y después chequear las respuestas.

   ```
   x * 2
   x * y
   x[1] * y[2]
   1/x
   (1:10) * x[2]
   rep(c(1,1,2), times = 2)
   ```

   b) ¿Qué ocurre si se multiplican 2 vectores de distinta longitud?

2. a) Crear un vector llamado tratamiento con coordenadas A, B y C de manera que A aparezca 20 veces, B 18 y C 22.

   b) Definir un vector J que sea una secuencia de 5 a 100 con un incremento de 5 y luego sumar la primera y la octava coordenadas.

3. Sea $q = (3, 0, 1, 6)$ y $r = (1, 0, 2, 4)$

   a) Calcular el producto escalar entre q y r
   b) Construir una matriz cuyas filas sean q y r. LLamarla a.
   c) Construir una matriz cuyas filas sean q y r. LLamarla b.
   d) Calcular el producto de a y b.
   e) Calcular la inversa de $a * b$.
   f) Reemplazar la primera columna de a por una columna de unos.
   g) Extraer la segunda columna de b.

4. Calcular la suma de los cuadrados de los números naturales del 1 al 100 usando R.

5. Considerar el conjunto de datos mtcars en R. (Ver `help(mtcars)`)

   a) ¿Qué autos tienen 4 velocidades?
   b) ¿Qué autos hay en el subconjunto `mtcars[mtcars$disp > 150 & mtcars$mpg > 20,]`?
   c) ¿Qué autos tienen 4 velocidades y transmisión manual?
   d) Hallar la cantidad media de millas por galón de los autos con 2 carburadores.

6. El conjunto de datos `arbolado-en-espacios-verdes.csv` contiene datos de 2011 de todos los árboles de los espacios verdes de la ciudad de Buenos Aires. Los datos pueden encontrarse aquí, junto con una descripción de los mismos .

a) ¿Cuántas filas tiene el conjunto de datos? ¿Cuántas columnas? ¿Cuáles son los nombres de las columnas?

b) Calcular la altura promedio de los árboles de la ciudad en 2011.

c) ¿Cuántos árboles había en 2011 en la plaza Arenales?

d) Construya un `data.frame` llamado `arboles_favoritos` que contenga sólo las filas correspondientes a su espacio verde favorito. Puede utilizar el comando `sort` para ordenar alfabéticamente un vector de caracteres.

e) Utilizando el comando `unique` averigue los nombres de los árboles presentes en su espacio verde favorito en 2011.

7. *a*) Definir una función llamada `elegir_alumnos` que elija al azar m alumnos de un aula de n para exponer ejercicios en el pizarrón, asumiendo que los alumnos estan numerados de 1 a n.

   b) Simular el experimento de elegir 3 alumnos al azar de un aula de 80 alumnos para exponer ejercicios en el pizarrón.

   c) Repetir 10 veces el experimento del item 7*b* y guardar los resultados en una matriz llamada `alumnos_elegidos`

8. Algunos dicen que las últimas horas a bordo del Titanic estuvieron marcadas por la guerra de clases, otros sostienen que estuvieron caracterizadas por la caballerosidad de los varones. Los datos TITANIC3 del paquete PASWR2 contienen información sobre los pasajeros del Titanic, incluyendo clase, sexo y si sobrevivieron o no, entre otras características.

   (a) Determine la proporción de sobrevivientes por clase.

   (b) Calcule la proporción de sobrevivientes por clase y sexo. ¿Quién tuvo una tasa más alta de supervencia: los varones de 1ra clase o las mujeres de 3ra?

   (c) Hacer un histograma con los datos correspondientes a la edad de los pasajeros con 20 intervos. ¿Qué observa? Si dividimos el rango de las edades en 20 grupos de 5 años cada uno: 0-5, 6-10, ..., 75-80. ¿Cuál era el rango de edades más numeroso en el Titanic? Y el menos numeroso? ¿Cómo trata R los valores faltantes al hacer el histograma?

   (d) ¿Cuál era la edad de la mujer más joven que sobrevivió?

   (g) En su opinión y basándose en estos datos ¿fue la guerra de clases, la caballerosidad de los varones o una combinación de ambas lo que caracterizó las últimas horas del Titanic?

9. El conjunto de datos CARS2004 del paquete PASWR2 contiene datos de automóviles en Europa del año 2004. (ver help(CARS2004))

   a) Calcular cantidad total de autos en cada país.

   b) Calcular la tasa de mortalidad de automovilistas para cada país como el número total de muertes de automovilistas dividido el número total de autos.

   c) Hacer un gráfico de barras que indique la tasa de mortalidad en accidentes de tránsito de cada país de la Unión Europea. Ordene las barras de forma creciente.

d) ¿Qué país tiene la menor tasa de mortalidad de automovilistas y qué país la más baja?

e) Haga un gráfico de cantidad total de autos vs población. ¿Cómo describiría la relación?

f) Haga un gráfico de cantidad total de autos vs tasa de mortalidad de automovilistas. ¿Cómo describiría la relación?

PROBABILIDADES Y ESTADÍSTICA (C)

Práctica 1

1. Se arroja dos veces un dado equilibrado, registrándose los resultados obtenidos.

   a) Definir un espacio muestral $\mathcal{S}$ apropiado para este experimento.

   b) Describir el conjunto de elementos del espacio muestral que satisface que:
   A: la suma de los dos números obtenidos es por lo menos 5
   B: el valor obtenido en el primer tiro es superior al obtenido en el segundo
   C: el valor obtenido en el primer tiro es un 4

   c) Calcular las probabilidades de los eventos definidos en 1*b)*.

   d) Simular en R el experimento de tirar dos veces un dado equilibrado.

   e) Simular 1000 veces en R el experimento de tirar dos veces un dado equilibrado y estimar las probabilidades de los sucesos definidos en 1*b)*.

   f) Describir los siguientes conjuntos:

      i. $A \cap B$
      ii. $B \cup C$
      iii. $A \cap (B \cup C)$

   g) Calcular las probabilidades de los sucesos definidos en 1*f)*.

   h) Estimar las probabilidades de los sucesos definidos en 1*f)* mediante simulaciones. Comparar con los resultados obtenidos en 1*g)*.

2. a) Dados dos eventos A y B tales que se conocen $P(A \cup B)$ y $P(A \cap B)$, hallar una fórmula para la probabilidad de que ocurra exactamente uno de estos eventos.

   b) Una compañía constructora trabaja en dos proyectos diferentes. Sea A el evento: “el primero de los proyectos se termina en la fecha del contrato” y definamos análogamente B para el segundo proyecto. Si $P(A \cup B) = 0.9$ y $P(A \cap B) = 0.5$, ¿cuál es la probabilidad de que exactamente un proyecto se termine para la fecha de contrato?

3. Supongamos que cuando una computadora se “cuelga” (no responde), el 75 % de las veces se debe a problemas de memoria y el 15 % de las veces a problemas de software y que el 15 % de las veces se debe a problemas que no son ni de memoria ni de software. Si una computadora se cuelga,

   a) ¿cuál es la probabilidad de que estos dos problemas ocurran simultáneamente?

   b) ¿cuál es la probabilidad de que ocurra un problema de software y no de memoria?

4. De un bolillero que contiene 5 bolillas numeradas 1, 2, 3, 4, 5 se extrae una al azar, sea la número k. Se eliminan las bolillas cuyo número es mayor que k de la urna y se hace una segunda extracción al azar entre las bolillas 1 a k, sea la número j. Se eliminan las bolillas cuyo número es mayor que j de la urna y se hace una tercera extracción al azar entre las bolillas 1 a j.

a) Describir un espacio muestral adecuado para este experimento y determinar el número de elementos que posee.

b) ¿Es razonable suponer equiprobabilidad en este espacio? ¿Qué probabilidad le asignaría al (3,2,1)?

5. Una firma proveedora de software ha ofrecido sus servicios a 3 empresas. Se definen los eventos $A_i = \{\text{la empresa } i \text{ realiza una compra a esta firma}\}$, para $i = 1, 2, 3$. Se sabe que

$$\begin{array}{lll} P(A_1) = 0.22 & P(A_2) = 0.25 & P(A_3) = 0.28 \\ P(A_1 \cap A_2) = 0.11 & P(A_1 \cap A_3) = 0.05 & P(A_2 \cap A_3) = 0.07 \\ P(A_1 \cap A_2 \cap A_3) = 0.01 & & \end{array}$$

Expresar en palabras los siguientes eventos y calcular sus probabilidades:

$$\begin{array}{ccc} A_1 \cup A_2 & A_1^c \cap A_2^c & A_1 \cup A_2 \cup A_3 \\ A_1^c \cap A_2^c \cap A_3^c & A_1^c \cap A_2^c \cap A_3 & (A_1^c \cap A_2^c) \cup A_3 \end{array}$$

6. Un grupo de 60 alumnos será subdividido al azar en dos divisiones de 30 alumnos cada una. Cinco de esos alumnos son muy amigos.: Alicia, Beto, Carmen, Diego y Eva.

a) ¿Cuál es la probabilidad de que todos queden en la misma división?

b) ¿Cuál es la probabilidad de que sólo quede separado Diego?

7. De un grupo de 6 mujeres y 4 hombres se deben elegir 3 personas para que los representen en tres congresos a desarrollarse en mayo, junio y septiembre.

a) Suponiendo que una persona puede ir a más de un congreso, calcular la probabilidad de que

i. a los dos primeros congresos vayan mujeres.

ii. a los dos primeros congresos vayan mujeres y al tercero un hombre.

iii. haya por lo menos una mujer entre las 3 personas elegidas.

b) Si a cada congreso debe ir una persona diferente, calcular las mismas probabilidades que en (a) y además la probabilidad de que haya exactamente una mujer entre las 3 personas elegidas.

8. En un negocio hay 6 productos de cierto tipo, 3 de ellos vencidos y 3 que están dentro del periodo de validez. Si un supervisor, que no sabe cuántos envases válidos o vencidos hay, los revisa, ¿cuál es la probabilidad de que

a) los tres primeros envases revisados contengan los productos vencidos?

b) necesite revisar exactamente i envases para encontrar los tres que contienen los productos vencidos? (hacerlo para $i = 4, 5, 6$).

9. a) Enunciar y probar una fórmula para $P(A_1 \cup A_2 \cup A_3 \cup A_4)$.

b) Cuatro matrimonios deciden bailar un tango, eligiendo las mujeres a sus compañeros al azar.

i. ¿Cuál es la probabilidad de que la mujer i (fijo) elija a su esposo como pareja de baile; i=1,2,3,4?

ii. ¿Cuál es la probabilidad de que al menos 1 mujer elija a su esposo?

a) Estimar las probabilidades del item b) por medio de una simulación. Compare con los resultados obtenidos en b).

10. Un estudio sobre la relación entre nivel de ingresos (A=alto, M=medio, B=bajo) y la preferencia por una de las tres grandes marcas de automóviles (Y,W,Z) da como resultado la siguiente tabla de probabilidades conjuntas.

	B	M	A	
Y	0.10	0.13	0.02	0.25
W	0.20	0.12	0.08	0.40
Z	0.10	0.15	0.10	0.35
	0.40	0.40	0.20	

Esta tabla muestra por ejemplo que:

$$\begin{array}{lclcl} P(\text{ingreso bajo y preferencia } Y) & = & P(B \cap Y) & = & 0.10 \\ P(\text{ingreso bajo}) & = & P(B) & = & 0.40 \\ P(\text{preferencia } Y) & = & P(Y) & = & 0.25 \end{array}$$

a) En base a esta tabla calcular las siguientes probabilidades condicionales:

$$\begin{array}{lll} P(W|A) & P(M|Z) & P(Y^c|M) \\ P(M|Y^c) & P(M|W \cup Z) & P(B \cup M|Z) \end{array}$$

b) ¿Cuál es la probabilidad de que una persona elegida al azar prefiera la marca Y o tenga un alto ingreso?

11. Una persona arroja dos dados equilibrados. Calcular la probabilidad de que la suma sea 7 dado que

a) la suma es impar.

b) la suma es mayor que 6.

c) el número del segundo dado es par.

d) el número de alguno de los dados es impar.

e) los números de los dados son iguales.

f) Estime las probabildades de los items anteriores por medio de una simulación.

12. Se lanzan 3 dados. Si ninguno muestra la misma cara, ¿cuál es la probabilidad de que haya salido exactamente un as?

13. Un sistema de computación on-line tiene 4 líneas de entrada para comunicación. Cada línea cubre un porcentaje del tráfico de entrada y cada línea tiene un porcentaje de mensajes que ingresan con error. La tabla a continuación describe estos porcentajes:

Línea	% de mensajes que entra por la Línea	% de mensajes sin error
1	40 %	99.8 %
2	30 %	99.9 %
3	10 %	99.7 %
4	20 %	99.2 %

a) ¿Cuál es la probabilidad de que al elegir al azar un mensaje, éste haya ingresado sin error?

b) ¿Cuál es la probabilidad de que, si el mensaje entró con error, haya ingresado por la línea 1?

14. Una caja contiene tres monedas, una de las cuales tiene dos caras, otra tiene dos cruces (o cecas) y la tercera es una moneda normal. Se extrae de la caja una moneda al azar y se arroja, obteniéndose una cara.

    a) ¿Cuál es la probabilidad de haber elegido la moneda con dos caras?

    b) Si se arroja nuevamente la moneda extraída, ¿cuál es la probabilidad de obtener otra vez cara?

15. Una enfermedad afecta a una de cada 500 personas de cierta población. Se usa un examen radiológico para detectar posibles enfermos. Se sabe que la probabilidad de que el examen aplicado a un enfermo lo muestre como tal es 0.90 y que la probabilidad de que el examen aplicado a una persona sana la muestre como enferma es 0.01.

    Calcular la probabilidad de que una persona esté realmente enferma si su examen dio positivo.

16. Hay tres cajas A, B y C con 20 piezas cada una, conteniendo 20, 15 y 10 piezas buenas respectivamente. La probabilidad de elegir la caja A es igual a la de elegir la caja B, y la de elegir la caja C es igual a la suma de esas dos probabilidades. Eligiendo al azar una caja se extraen con reposición dos piezas que resultan ser buenas. Hallar la probabilidad condicional de que provengan de la caja A.

17. ($*$) Se dispone de dos urnas, cuyo contenido es el siguiente:

    Urna A: 5 bolitas rojas y 3 blancas.

    Urna B: 1 bolita roja y 2 blancas.

    Se arroja un dado equilibrado. Si sale 3 ó 6 se extrae una bolita de la urna A y se la coloca en B, luego se saca una bolita de B. En caso contrario el proceso se hace a la inversa.

    a) Hallar la probabilidad de que ambas bolitas sean rojas.

    b) Si ambas bolitas son rojas, ¿cuál es la probabilidad de que haya salido 3 ó 6?

18. Se tienen $(n+1)$ urnas numeradas $0, 1, \ldots, n$. La urna i contiene i bolitas blancas y $(n-i)$ bolitas negras. Se elige al azar una urna y de ella se extrae una bolita al azar. Sugerencua

    a) Calcular la probabilidad de que la bolita extraída sea blanca.

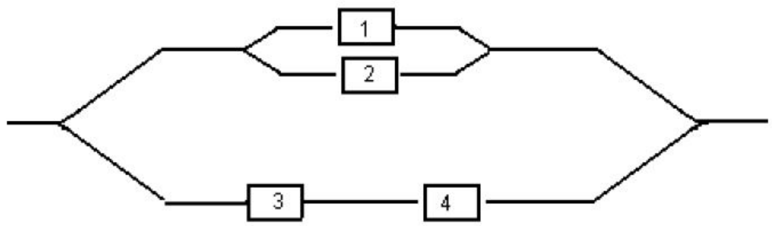

Figura 1:

b) Si la bolita extraída es blanca, calcular la probabilidad de que provenga de la urna i, $(i = 0, 1, \ldots, n)$. Sugerencia: usar que $\sum_{i=1}^{n} i = n(n+1)/2$.

19. Se extrae al azar una bolita de una urna que contiene 9 bolitas de las cuales 3 son blancas, 3 son negras y 3 son rojas. Las bolitas están numeradas 1, 2 y 3 dentro de cada color y además las siguientes bolitas están rayadas: número uno color blanco, número dos color negro y número tres color rojo. Sean los sucesos:

    A: la bolita es número uno.

    B: la bolita es blanca.

    C: la bolita es rayada.

    a) ¿Son independientes de a pares los sucesos A, B y C?
    b) ¿Son independientes los sucesos A, B y C?

20. Considérese un sistema de componentes conectados como lo muestra la Figura 1. Los componentes 1 y 2 están conectados en paralelo, de manera que el subsistema funciona si y sólo si cualquiera de ellos funciona; en cambio los componentes 3 y 4 están conectados en serie y por lo tanto este subsistema funciona si y sólo si ambos funcionan. El sistema funciona si al menos uno de los dos subsistemas funciona.

    Si los componentes trabajan independientemente y la probabilidad de que un componente cualquiera funcione es 0.9,

    a) calcular la probabilidad de que el sistema funcione.
    b) calcular la probabilidad de que el componente 1 no funcione si se sabe que el sistema sí funciona.

21. (*) Sean $S_1, \ldots, S_n$ sucesos independientes tales que $P(S_i) = p_i$.

    a) Expresar en función de los p_i la probabilidad de que de los S_i ocurra
        i. ninguno.
        ii. al menos uno.
        iii. exactamente uno.
    b) Para $k = 0, \ldots, n$, hallar la probabilidad de que ocurran exactamente k de los sucesos S_i en el caso en que $P(S_i) = p$ para todo i.

22. ($*$) Sean A y B dos eventos de un espacio muestral Ω.

a) Probar que el suceso A es independiente de cualquier suceso B si y sólo si P(A)=0 ó P(A)=1.

b) Probar que si A $\subset$ B y A y B son independientes, entonces P(A)=0 ó P(B)=1.

PROBABILIDADES Y ESTADÍSTICA (C)

Práctica 2

1. De un lote que contiene 15 artículos, de los cuales 4 son defectuosos, se eligen 3 artículos al azar con reposición. Si llamamos X al número de artículos defectuosos entre los seleccionados,

   a) Hallar la función de probabilidad puntual asociada a X y graficarla usando R.

   b) ¿Cuál es la probabilidad de que al menos 2 artículos sean defectuosos?

   c) Hallar la función de distribución acumulada de X y graficarla usando R.

   d) Estimar mediante una simulación las probabilidades calculadas en el item a).

2. Sea X una variable aleatoria con función de distribución

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ 0.3 & \text{si } 1 \leq x < 3 \\ 0.4 & \text{si } 3 \leq x < 6 \\ 0.6 & \text{si } 6 \leq x < 12 \\ 1 & \text{si } 12 \leq x \end{cases}$$

   a) Hallar la función de probabilidad puntual de X.

   b) Calcular de dos maneras: utilizando la función de distribución y utilizando la función de probabilidad puntual, las siguientes probabilidades:

$$P(3 < X \leq 6) \qquad P(3 \leq X \leq 6) \qquad P(X \geq 4) \qquad P(X \geq 6)$$

   c) Utilizando el comando `sample`, generar 5 realizaciones de esta variable aleatoria en R.

   d) Mediante una simulación, estimar las probabilidades del item b).

3. Definir una función en R que, dados dos vectores $x = (x_1, \ldots, x_n)$ y $p = (p_1, \ldots, p_n)$ y un número real t, calcule $F_X(t)$, es decir, la probabilidad de que de la variable aleatoria discreta que toma valores $x_1, \ldots, x_n$ con probabilidades $p_1, \ldots, p_n$ respectivamente, sea menor o igual que t. Probarla con la variable aleatoria del ejercicio 2.

4. Si X es una v.a. discreta que toma sólo valores enteros, probar que para todo $k \in Z$:

$$p_X(k) = F_X(k) - F_X(k-1)$$

5. Calcular la esperanza de la variable aleatoria definida en el ejercicio 1 utilizando la definición y estimarla usando una simulación. Comparar los resultados.

6. Definir una función en R que, dados dos vectores $x = (x_1, \ldots, x_n)$ y $p = (p_1, \ldots, p_n)$ calcule la esperanza de la variable aleatoria discreta que toma valores $x_1, \ldots, x_n$ con probabilidades $p_1, \ldots, p_n$ respectivamente. Probar que funciona para la variable aleatoria X del ejercicio 1.

7. Definir una función en R que, dados dos vectores $x = (x_1, \dots, x_n)$ y $p = (p_1, \dots, p_n)$ calcule la varianza de la variable aleatoria discreta que toma valores $x_1, \dots, x_n$ con probabilidades $p_1, \dots, p_n$ respectivamente. Probar que funciona para la variable aleatoria X del ejercicio 1.

8. Sea X una v.a. con distribución Bernoulli de parámetro p.

   a) Calcular $E(X^k)$ para $k \in \mathbb{N}$
   b) Mostrar que $V(X) = p(1-p)$.

9. El 70% de las consultas de un sistema interactivo de computación requiere de acceso a bases de datos. Un sistema recibe 25 consultas independientes unas de otras,

   a) ¿cuál es la probabilidad de que:
      i. exactamente 20 consultas requieran acceso a una base de datos?
      ii. el número de consultas que requieran acceso a una base de datos esté entre 20 y 24 inclusive?
   b) Calcular el valor esperado y la varianza del número de consultas que requieren acceso a una base de datos.

10. Se tienen dos dados, uno equilibrado y el otro cargado en el cual los números 1 y 2 tienen probabilidad 1/3 y el resto 1/12. Se elige un dado al azar y se lo arroja tres veces (independientemente). Sea X el número de veces que sale 1 ó 2.

    a) ¿Cuál es la distribución de X condicional a que se eligió el dado cargado?
    b) Hallar una expresión general para $p_X(k)$.

11. Para verificar si se cumplen las normas establecidas para arrojar residuos al río Reconquista, un inspector visita al azar 10 de las 50 industrias establecidas a orillas de dicho río.

    a) Si en realidad 35 industrias no cumplen con alguna de las normas, ¿cuál es la distribución del número de industrias visitadas que están en infracción? Calcular la probabilidad de que 6 de las industrias visitadas estén en infracción.
    b) Si hay 500 industrias de las cuales 350 están en infracción, aproximar la distribución de (a) por una más simple. Calcular nuevamente la probabilidad de que 6 de las industrias visitadas estén en infracción.
    c) Sea X el número de fábricas que están en infracción entre las 10 visitadas. Calcular $E(X)$ y $V(X)$ para las distribuciones exacta (a) y aproximada (b).

12. Una rueda de ruleta está dividida en 38 secciones, de las cuales 18 son rojas, 18 son negras y las 2 restantes son verdes. Sea X el número necesario de juegos hasta obtener una sección verde en jugadas independientes.

    a) ¿Cuál es la probabilidad de que sean necesarias al menos 4 jugadas?
    b) Hallar la función de distribución acumulada de la v.a. X.

c) Si fueron necesarias 7 o más jugadas, ¿cuál es la probabilidad de que se necesiten al menos 10 jugadas? Comparar con (a).

d) ¿Cuál es la probabilidad de que sea necesario un número impar de jugadas?

e) Hallar $E(X)$ y $V(X)$.

13. Si en el ejercicio anterior se define Y: número de juegos hasta obtener exactamente tres secciones verdes,

    a) ¿qué distribución tiene la v.a. Y?

    b) ¿cuál es la probabilidad de que se requieran exactamente 5 jugadas?

    c) hallar $E(Y)$ y $V(Y)$.

14. Con el fin de encontrar una palabra clave, un motor de búsqueda de internet explora una secuencia de sitios de la WEB en orden aleatorio. Al iniciar la búsqueda, el motor elige, al azar y con igual probabilidad, una entre dos secuencias posibles de sitios. Se sabe que el 10% de los sitios de la primera secuencia contienen esta palabra clave, mientras que sólo el 5% de los sitios de la segunda contienen dicha palabra.

    a) Si la búsqueda termina ni bien se encuentra un sitio que contenga la palabra clave, ¿cuál es la probabilidad de que más de 5 sitios deban ser explorados?

    b) Si se sabe que el motor de búsqueda encontró la palabra clave en la sexta visita ¿cuál es la probabilidad de que la haya encontrado en la segunda secuencia?

    c) Si la búsqueda termina cuando se encuentran 2 sitios que contenga la palabra clave ¿cuál es la probabilidad de que deban explorarse exactamente 10 sitios?

15. Un minorista ha verificado que la demanda semanal de cajones de cierto producto es una v.a. con distribución de Poisson de parámetro $\lambda = 2$. Completa su existencia los lunes por la mañana de manera de tener 4 cajones al principio de la semana. Al efectuar un análisis de la actividad de su negocio, se le plantean las siguientes preguntas:

    a) ¿Cuál es la probabilidad de vender todo su stock durante la semana?

    b) ¿Cuál es la probabilidad de que sea incapaz de cumplir con un pedido por lo menos?

    c) ¿Cuál es la distribución del número de cajones vendidos en una semana?

    d) ¿Con cuántos cajones debería iniciar la semana a fin de que la probabilidad de cumplir con todos sus pedidos fuese mayor o igual que 0.99?

16. Un bibliotecario ubica 1000 libros en un cierto día. Si la probabilidad de que un libro cualquiera sea mal ubicado es 0.001 y los libros se ubican en forma independiente, ¿cuál es la distribución aproximada del número de libros mal ubicados en ese día?

    Utilizando esta distribución, calcular la probabilidad de que

    a) por lo menos un libro sea mal ubicado ese día.

    b) exactamente 3 libros sean mal ubicados ese día. Comparar con el valor exacto.

17. En un concurso de pesca cada pescador paga 100$ por participar. La cantidad de peces obtenida por cada pescador durante el desarrollo del concurso es una v.a. con distribución de Poisson de parámetro $\lambda = 4$. Hay un premio de 50$ por pieza. Cada pescador tiene permitido cobrar a lo sumo 6 piezas (es decir, aunque pesque más de 6 cobrará sólo por 6).

    a) Calcular la función de probabilidad puntual de la ganancia neta de un pescador.
    b) ¿Cuánto dinero espera ganar cada participante?

18. Sea X una variable aleatoria binomial con $n = 5$ y $p = 1/3$.

    (a) Graficar la función de probabilidad puntual p_X y la función de distribución acumulada F_X, basándose en los comandos `dbinom` y `pbinom` respectivamente.
    (b) Graficar la función de distribución acumulada de X utilizando la función implementada en el ejercicio 3. Comparar con el gráfico obtenido en (a).
    (c) Utilizando el comando `rbinom` generar 1000 realizaciones de la variable aleatoria X y calcular la proporción de veces que X toma los valores 0, 1, 2, 3, 4 y 5. Comparar con los valores que toma la fpp.

19. Las tareas llegan a una cola de un sistema de computación con un solo servidor de acuerdo con un proceso de Poisson de parámetro $\lambda = 5$ tareas por minuto. Llamemos a dicho proceso X_t

    a) ¿Cuál es la probabilidad de que en 30 segundos lleguen menos de 5 tareas?
    b) Sea X_1 la cantidad de tareas recibidas en un minuto. Calcular:

    $$\begin{array}{ll} P(X_1 \leq 6) & P(3 \leq X_1 \leq 6) \\ P(X_1 = 6) & P(3 < X_1 < 6) \\ P(X_1 \geq 5) & P(3 \leq X_1 \leq 6 | X_1 \geq 4) \end{array}$$

    Calcularlas exactamente utilizando los comandos `dpois` o `ppois` y luego estimarlas mediante una simulación (utilizar el comando `rpois`).
    c) ¿Cuál es el número esperado de tareas que se reciben en media hora?
    d) Estimar la esperanza calculada en el item anterior usando R.

20. El número de veces que una red de computadoras se bloquea sigue un proceso de Poisson de parámetro igual a 2 bloqueos por semana. Hallar la probabilidad de que

    a) en 2 semanas no se bloquee.
    b) en un periodo de 4 semanas, haya exactamente 1 semana en la que no se bloquea. (Sugerencia: Notar que se está preguntando por cantidad de semanas, no de bloqueos.)

PROBABILIDADES Y ESTADÍSTICA (C)

Práctica 3

1. Sea X una v.a. con función de densidad

$$f_X(x) = \begin{cases} 0{,}75\,(1-x^2) & -1 \leq x \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

   a) Verificar que f_X es realmente una función de densidad.

   b) Calcular:

$$P(X > 0) \qquad P(-0{,}5 < X < 0{,}5) \qquad P(|X| > 0{,}25)$$

   c) Hallar F_X, la función de distribución acumulada de X.

   d) Graficar f_X y F_X.

2. Sea X una v.a. continua con función de distribución

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{x^3}{\theta} & \text{si } 0 \leq x < 2 \\ 1 & \text{si } 2 \leq x \end{cases}$$

   a) ¿Cuál es el valor de θ?

   b) Calcular, usando $F_X(x)$,

$$P(X \leq 1) \qquad P(0{,}5 \leq X \leq 1) \qquad P(0{,}5 < X \leq 1 | X < 1)$$

   c) Hallar la mediana $\tilde{\mu}$ de esta distribución.

   d) Encontrar la función de densidad $f_X(x)$.

   d) Graficar f_X y F_X. Agregar la recta vertical $x = \tilde{\mu}$.

3. Sea U una v.a. con distribución $\mathcal{U}[0, \theta]$, para $\theta > 0$.

   a) Hallar la función de distribución acumulada de U.

   b) Si $P(1 \leq U \leq 3) = 0{,}5$, ¿qué valores puede tomar θ? para

   c) Para θ hallados en el item b) estimar la $P(U^2 < 2)$.

4. Consideremos una v.a. Y con función de densidad

$$f_Y(y) = \begin{cases} \dfrac{1}{25}y & 0 \leq y < 5 \\ \dfrac{2}{5} - \dfrac{1}{25}y & 5 \leq y < 10 \\ 0 & \text{en otro caso.} \end{cases}$$

a) Calcular la función de distribución de Y.

b) Calcular $E(Y)$ y $V(Y)$.

c) Calcular $E(1/Y)$. ¿Qué conclusión saca respecto a la relación entre $E(1/Y)$ y $1/E(Y)$.

5. Se eligen n puntos al azar en el intervalo $[0,1]$ de forma independiente (con distribución uniforme). Sea X = Cantidad de puntos que caen en el intervalo $[0,p]$ $(0 < p < 1)$. ¿Qué distribución tiene X?

6. Sea Z una v.a. con distribución $N(0,1)$. Calcular:

   a) $P(0 \leq Z \leq 2)$

   b) $P(|Z| \leq 2{,}5)$

   c) $P(Z \geq -1{,}37)$

   d) c tal que $P(Z < c) = 0{,}98$

   e) c tal que $P(|Z| \leq c) = 0{,}90$

   f) el valor z_α para $\alpha = 0{,}1, 0{,}05, 0{,}025, 0{,}01$, donde z_α se define como el valor tal que $P(Z > z_\alpha) = \alpha$

7. Sea X una v.a. con distribución $N(5, 0{,}25)$. Calcular:

   a) $P(4{,}75 \leq X \leq 5{,}50)$

   b) $P(|X| > 5{,}25)$

   c) c tal que $P(|X - 5| \leq c) = 0{,}90$

   d) el 90-percentil de X.

8. Se supone que en cierta población humana, el índice cefálic I (anchura del cráneo expresada como porcentaje de la longitud) es una v.a. con distribución $N(\mu, \sigma^2)$. Si hay un 58 % de individuos con $I \leq 75$, un 38 % con $75 < I \leq 80$ y un 4 % con $I \geq 80$, hallar la función de densidad de I y calcular $P(78 \leq I \leq 82)$.

9. La distancia intercuartil (IQR) de una v.a. se define como la diferencia entre el tercer cuartil y el primer cuartil. Hallar la IQR de Z, una v.a. con distribución $N(0,1)$.

10. (*) La mediana de desviaciones absolutas (MAD) de una v.a. se define como la mediana del valor absoluto de la diferencia entre la variable aleatoria y su mediana. Hallar la MAD de Z, una v.a. con distribución $N(0,1)$. *Sugerencia:* Hallar la f.d.a de $|Z|$.

11. La biblioteca de una facultad dispone de una red de computadoras al alcance de los estudiantes. La proporción de tiempo que un usuario destina a búsqueda bibliográfica es una variable aleatoria T con función de densidad

$$f_T(t) = c\,(100 - t)\, I_{[0,100]}(t)$$

   a) Hallar el valor de la constante c

b) Supóngase que de acuerdo con el porcentaje de tiempo destinado a las búsqueda bibliográfica el usuario es clasificado en distintas categorías. La clasificación se realiza de la siguiente manera: si $T < 25\,\%$ el usuario es de tipo 1, si $25\,\% \leq T < 50\,\%$ el usuario es tipo 2, si $50\,\% \leq T < 75\,\%$ el usuario es tipo 3 y si $T > 75\,\%$, el usuario es de tipo es 4. Hallar la distribución del tipo de usuario.

12. El diámetro D (expresado en dm) del tronco de cierta especie de árboles es una variable aleatoria con función de densidad

$$f_D(x) = k\,x\,I_{(0,10)}(x).$$

a) Hallar el valor de la constante k.

b) ¿Cuál es la probabilidad de que el diámetro de un árbol de esa especie elegido al azar mida entre 4 y 6 dm?

c) Idem b) sabiendo que el diámetro mide más de 5 dm.

d) En un área del bosque hay 3 árboles de esa especie. Calcular la probabilidad de que exactamente 2 de ellos tengan el diámetro entre 4 y 6 dm.

e) ¿Cuántos árboles habría que muestrear en el bosque para que la probabilidad de encontrar al menos uno cuyo diámetro mida entre 4 y 6 dm, sea mayor o igual que 0.99?

13. Una muestra de tabaco puede provenir de dos variedades distintas, I y II, con probabilidades 0.35 y 0.65 respectivamente. El contenido de nicotina es una v.a., cuya distribución es $N\,(1{,}9, 0{,}16)$ en la variedad I y $N\,(2{,}2, 0{,}09)$ en la variedad II.

a) Hallar la probabilidad de que en una muestra elegida al azar el contenido de nicotina sea mayor o igual que 2.1.

b) Hallar la probabilidad de que dado que el contenido de nicotina es mayor que 2.1, la muestra provenga de la variedad I.

14. La porción de memoria ocupada en un servidor de un sistema de terminales en red es una variable aleatoria continua X que toma valores entre 0 (sin carga) y 1 (carga completa). La densidad de X está dada por

$$f_X(x) = 4x^3 \quad \text{si} \quad 0 < x < 1$$

a) Halle la mediana de la porción ocupada de memoria.

b) Deduzca la densidad de la variable que mide la porción de memoria que falta ocupar, es decir $Z = 1 - X$.

15. El tiempo de caída de un sistema se define como la fracción de tiempo que el sistema no está operativo debido a una falla del hardware o del software. Supongamos que T= tiempo de caída de un sistema en horas es una variable aleatoria con función de densidad dada por

$$te^{-\frac{t^2}{2}} \quad I_{(0,\infty)^{(x)}}$$

a) Deduzca la función de distribución acumulada de T.

b) Cuando el sistema está caído por más de una hora, todos los archivos de trabajo abiertos en el momento de la caída se pierden. Si un usuario está trabajando en un archivo mientras el sistema cae, ¿cuál es la probabilidad de que el archivo no se pierda?

c) Supongamos que al caer el sistema, el tiempo que tarda un usuario en recuperar su trabajo es una función creciente del tiempo de caída, digamos T^2. Deduzca la función de densidad de T^2

16. La vida útil (en meses) de un componente electrónico es una v.a. V con distribución $\mathcal{E}(\lambda)$, tal que $P(V > 20) = 0{,}449$.

    a) Hallar $E(V)$ y $Var(V)$.

    b) Hallar la probabilidad de que la vida útil de uno de estos componentes sea mayor que 10 meses.

    c) Si se sabe que uno de estos componentes dura más de 20 meses, ¿cuál es la probabilidad de que dure más de 30 meses? Comparar con (b).

    d) Si se sabe que uno de estos componentes dura más de t meses, ¿cuál es la probabilidad de que dure más de $t + s$ meses ($t > 0$, $s > 0$)?

17. Las tareas llegan a una cola de un sistema de computación con un solo servidor de acuerdo con un proceso de Poisson de parámetro $\lambda = 4$ tareas por minuto. Llamemos a dicho proceso X_t. Sea T el tiempo de espera hasta que llegue la primera tarea medido en minutos. Calcule la probabilidad de que a lo sumo haya que esperar 15 segundos hasta que arribe la primera tarea.

18. Un sistema consta de 5 componentes electrónicos como los del Ejercicio 14, conectados en serie. Al fallar uno cualquiera de éstos, automáticamente se desconecta el sistema. Se supone que los tiempos de vida de los componentes son independientes. Es decir, si se definen los eventos

$$A_i = \{\text{el i-ésimo componente dura por lo menos hasta el instante } t\},$$

con $i = 1, \ldots, 5$, estos eventos son independientes.

Sea X el momento en el cual el sistema se desconecta.

    a) Escribir el evento $\{X \geq t\}$ en función de los A_i.

    b) Usar la independencia de los A_i para calcular $P(X \geq t)$.

    c) Hallar las funciones de distribución y de densidad de X. ¿A qué familia pertenece esta distribución?

19. Un programa tiene tres bloques. El tiempo de compilación en segundos de cada bloque es una variables aleatoria exponencial λ =1 y es independiente del tiempo de compilación de los otros bloques.

a) Calcule la probabilidad de que alguno de los tres bloques tenga un tiempo de compilación superior a los 2 segundos.

b) ¿Cuál es la probabilidad de que exactamente dos bloques tengan un tiempo de compilación superior a los 2 segundos?

20. Definir una función en R que, dada una función de densidad f y su soporte S, calcule la esperanza y la varianza de la variable aleatoria continua que tiene densidad f (con soporte S). Probar que funciona para la variables aleatoria X del ejercicio 1 de esta práctica. *Sugerencia:* Usar el comando `integrate`.

21. Sea X una v.a. con función de densidad $\Gamma(\alpha, \lambda)$:

$$f(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} I_{[0,\infty)}(x), \quad \text{con } \alpha > 0, \lambda > 0.$$

   a) Hallar $E(X)$ y $V(X)$.
   (Nota: Usar que $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ para $\alpha > 0$.)

   b) Probar que si $X \sim \Gamma(\alpha, \lambda)$ y $c > 0$ entonces $cX \sim \Gamma(\alpha, \lambda/c)$.

22. Sea X una v.a. con distribución $U(0, 1)$. Hallar las funciones de distribución y de densidad de las siguientes variables aleatorias.

   a) $cX + d$

   b) X^α, para $\alpha = 2$ y $\alpha = 3$

   c) $\ln X$

   d) $-\frac{1}{\lambda}\ln(1 - X)$, siendo $\lambda > 0$.

23. Sea Z una v.a. con distribución normal standard. Pruebe que $Z^2 \sim \Gamma(1/2, 1/2)$. (Esta variable aleatoria recibe el nombre de χ^2 con un grado de libertad).
   (Nota: $\Gamma(1/2) = \sqrt{\pi}$.)

24. (*) Sea X una v.a. con función de distribución $\mathcal{E}(2)$. Se define $Y = [X] + 1$ (donde $[\cdot]$ es la parte entera). Probar que $Y \sim \mathcal{G}(1 - e^{-2})$.

25. **Generación de números al azar**

   a) Usando el Ejercicio 22 parte a), generar, a partir de una variable aleatoria con distribución $U(0, 1)$, una variable aleatoria con distribución $U(3, 8)$.

   b) Usando el Ejercicio 19 parte d), generar, a partir de una variable aleatoria con distribución $U(0, 1)$, una variable aleatoria con distribución exponencial de parámetro 10.

   c) Generar, a partir de una variable aleatoria con distribución $U(0, 1)$, una variable aleatoria con la siguiente distribución uniforme discreta:

$$p_X(k) = \frac{1}{100} \qquad \text{si } k = 0, 1, ..., 99.$$

d) Generar, a partir de una variable aleatoria con distribución $U(0,1)$, una variable aleatoria con distribución $Bi(1,1/3)$.

e) Generar, a partir de una variable aleatoria con distribución $U(0,1)$, una variable aleatoria con distribución de Poisson de parámetro 5.

PROBABILIDADES Y ESTADÍSTICA (C)

Práctica 4

1. Sea (X, Y) un vector aleatorio con función de densidad conjunta

$$f_{XY}(x,y) = \begin{cases} k\left(x^2+y^2\right) & 2 \leq x \leq 3,\ 2 \leq y \leq 3 \\ 0 & \text{en otro caso.} \end{cases}$$

   a) ¿Cuál es el valor de k?
   b) ¿Cuál es la probabilidad de que tanto X como Y sean menores que 2,6?
   c) ¿Cuál es la probabilidad de que $\max(X, Y) < 2{,}6$?
   d) Hallar f_X y f_Y, las funciones de densidad marginales.

2. De un grupo de tres profesores, dos graduados y un alumno debe seleccionarse al azar una comisión de dos personas. Sean X el número de profesores e Y el número de graduados en la comisión.

   a) Hallar la función de probabilidad conjunta del par (X, Y) y las marginales de X e Y.
   b) ¿Cuál es la probabilidad de que el alumno no forme parte de la comisión?

3. Sea (X, Y) una v.a. bidimensional continua con distribución uniforme en el trapecio de vértices $(-1, 0)$, $(0, 1)$, $(1, 1)$ y $(2, 0)$.

   a) Hallar la función de densidad conjunta de (X, Y).
   b) Calcular $P(Y \leq X)$.
   c) Hallar las funciones de densidad marginales f_X y f_Y.

4. Para iluminar sin interrupción una sala (de computadoras!) se cuenta con dos lamparitas; cuando se quema una, se coloca la otra. Sean X e Y los tiempos de vida de cada lamparita (en 10^3 horas) y supóngase que esos tiempos son independientes y tienen distribución $\mathcal{E}(1)$.

   a) Hallar la densidad conjunta de (X, Y).
   b) Hallar la probabilidad de que la sala permanezca iluminada al menos 2000 horas.

5. Dos servidores A y B procesan trabajos a medida que van llegando. El tiempo que tarda el servidor A en procesar un trabajo es una variable aleatoria $X \sim \mathcal{E}(\lambda_1)$, mientras que el tiempo que tarda el servidor B es una variable aleatoria $Y \sim \mathcal{E}(\lambda_2)$. Ambos servidores actúan en forma independiente. Dos trabajos llegan simultáneamente y son atendidos uno por A y otro B. ¿Cuál es la probabilidad de que el servidor A termine con su trabajo antes que B?

6. (*) Retomemos el problema anterior y supongamos que tres trabajos llegan simultáneamente. Uno es atendido por A, otro por B y el tercero queda esperando a que uno de los servidores se libere. Si $\lambda_1 = \lambda_2$, pruebe que la probabilidad de que el último trabajo en ser atendido sea el último en ser completado es 0.5.

7. En los ejercicios 1 a 3:

   a) Decir si X e Y son independientes (justificando en cada caso).

   b) Hallar las funciones de probabilidad puntual condicional o de densidad condicional, según corresponda.

8. Se tira una moneda equilibrada 3 veces, siendo X el número de caras. Si $X = a$ se extraen sin reposición $a + 1$ bolillas de una urna que contiene 4 bolillas blancas y una roja. Sea Y el número de bolillas rojas extraídas.

   a) Hallar la distribución de Y dado $X = a$, para $a = 0, 1, 2, 3$.

   b) Obtener una tabla con la distribución conjunta del par (X, Y) y hallar la función de probabilidad puntual marginal p_Y.

   c) ¿Son X e Y independientes?

   d) Si se extrajeron 2 bolillas blancas, ¿cuál es la probabilidad de que hayan salido dos caras?

9. (*) Alicia y José acordaron encontrarse a las 8 de la noche para ir al cine. Como no son puntuales, se puede suponer que los tiempos X e Y en que cada uno de ellos llega son variables aleatorias con distribución uniforme entre las 8 y las 9. Además se supondrá que estos tiempos son independientes.

   a) ¿Cuál es la densidad conjunta de X e Y?

   b) ¿Cuál es la probabilidad de que ambos lleguen entre las 20:15 y las 20:45?

   c) Si ambos están dispuestos a esperar no más de 10 minutos al otro a partir del instante en que llegan, ¿cuál es la probabilidad de que se desencuentren?

10. En los ejercicios 1 a 3, calcular:

    a) $\operatorname{cov}(X, Y)$ y $\rho(X, Y)$.

    b) $E(X + Y)$ y $V(X + Y)$.

11. Implementar una función en R que, dada la tabla de la función de probabilidad puntual conjunta de un vector aleatorio (X, Y), devuelva la covarianza y la correlación entre X e Y.

12. (*) Sea (X, Y) un vector aleatorio con distribución uniforme en la región: $0 \leq x \leq 1, \quad x \leq y \leq x + h$ para todo $0 < h < 1$.

    (a) Calcular $E(X), E(Y), E(XY)$.

    (b) Hallar ρ_{XY}.

    (c) ¿A qué tiende ρ_{XY} cuando h tiende a cero? ¿Por qué?

13. (*) Sea X una variable aleatoria con densidad $\mathcal{U}[0, 1]$. Si $X = x$, se elige un número Y entre 0 y x. Por lo tanto $Y|_{X=x} \sim \mathcal{U}[0, x]$.

a) Hallar la densidad conjunta del par (X, Y) y la densidad marginal f_Y.

b) Calcular $E(Y)$, $V(Y)$, $\operatorname{cov}(X, Y)$ y $\operatorname{cov}(X, X+Y)$.

14. Se va a guardar un archivo de longitud 100 caracteres, cada uno de los cuales toma el valor A, B, C ó D. Las probabilidades de ocurrencia de cada uno de estos caracteres son $p_A = 0{,}70$, $p_B = 0{,}12$, $p_C = 0{,}10$ y $p_D = 0{,}08$. Se definen:

A: número de veces que ocurre la letra A.

B: número de veces que ocurre la letra B.

etc. Suponiendo independencia:

a) ¿Qué distribución tiene el vector aleatorio (A, B, C, D)?

b) Hallar las distribuciones marginales de A, B, C y D.

c) Para ahorrar memoria se decide representar estos caracteres según la siguiente tabla basada en el código de Huffman.

Letra	Código
A	1
B	00
C	011
D	010

Sea X = Tamaño del archivo codificado (en bits).

d) Hallar $E(X)$.

15. La longitud de ciertas varillas de metal tiene distribución $N(5,\ 0{,}25)$. Para hacer un control de calidad se eligen 40 varillas al azar. Hallar la probabilidad de que 1 varilla mida menos de 4 cm, 13 varillas midan entre 4 y 4.8 cm, 18 varillas midan entre 4.8 y 5.5 cm, y el resto de las varillas mida más de 5.5 cm.

16. Sean $X_1, \ldots, X_n$ v.a. i.i.d. (variables aleatorias independientes e idénticamente distribuídas) con función de distribución F_X. Se definen las variables aleatorias

$$\begin{aligned} T &= \max(X_1, \ldots, X_n) \\ U &= \min(X_1, \ldots, X_n) \end{aligned}$$

a) Probar que $F_T(t) = [F_X(t)]^n$.

b) Probar que $F_U(t) = 1 - [1 - F_X(t)]^n$.

c) Si las variables X_i tienen densidad $f_X(x)$, hallar las densidades de U y T en fución de $f_X(x)$ y $F_X(x)$.

17. En un concurso de salto, cada atleta salta 3 veces, siendo su puntaje la máxima altura alcanzada. Si la altura en cada salto (en metros) es una v.a. con distribución $\mathcal{U}[2, 2{,}5]$ y considerando que las tres alturas alcanzadas son independientes:

a) Hallar la función de distribución del puntaje.

b) Hallar el valor esperado del puntaje.

18. (*) Sean X e Y v.a. independientes, tales que $X \sim Bi(n,p)$ e $Y \sim Bi(m,p)$. Probar que:

    a) $X+Y \sim Bi(n+m,p)$.
    (NOTA: $\sum_{j=0}^{k} \binom{m}{k-j}\binom{n}{j} = \binom{m+n}{k}$.)

    b) La distribución de X condicional a $X+Y=k$ es $\mathcal{H}(k,n,m+n)$.

19. Un laboratorio posee 14 ratas de la especie A y 11 de la especie B para experimentación. La probabilidad de que cualquiera de ellas muera en un experimento es 0.1, y se considera que las ratas mueren en forma independiente.

    a) ¿Cuál es la probabilidad de que mueran más de 4 ratas?

    b) Si murieron 2, ¿cuál es la probabilidad de que ambas hayan sido de la especie A?

20. (*) Sean X e Y v.a. independientes, ambas con distribución $\mathcal{G}(p)$. Probar que $X+Y \sim BN(2,p)$.

21. (*) Sean $X_1,\ldots,X_n$ v.a. i.i.d.. Se define $S=\sum_{i=1}^{n} X_i$.

    a) Calcular $E(S)$ y $V(S)$ para los siguientes casos:
    i. $X_i \sim Bi(1,p)$.
    ii. $X_i \sim \mathcal{G}(p)$.

    b) Usando los Ejercicios 18 y 20, hallar la distribución de S para los dos casos anteriores.

    c) Deducir de (a) y (b) la esperanza y la varianza de variables aleatorias con distribución $Bi(n,p)$ y $BN(r,p)$.

22. (*) Sean X e Y v.a. independientes tales que $X \sim \mathcal{P}(\lambda)$ e $Y \sim \mathcal{P}(\mu)$. Probar que:

    a) $X+Y \sim \mathcal{P}(\lambda+\mu)$.

    b) La distribución de X condicional a $X+Y=k$ es $Bi\left(k, \dfrac{\lambda}{\lambda+\mu}\right)$.

    c) Sean X e Y v.a. tales que $X \sim \mathcal{P}(\lambda)$ e $Y|_{X=k} \sim Bi(k,p)$. Probar que $Y \sim \mathcal{P}(\lambda p)$.

23. Dos terminales A y B están conectadas a un servidor. La cantidad de requerimientos que realiza la terminal A en el lapso de un segundo sigue una distribución $\mathcal{P}(2)$ mientras que para la terminal B sigue una distribución $\mathcal{P}(3)$. Ambas terminales actúan en forma independiente.

    a) Hallar la probabilidad de que en un segundo haya más de 3 requerimientos al servidor.

b) Si en un determinado segundo hubo dos requerimientos al servidor, ¿cuál es la probabilidad de que haya provenido uno de cada terminal?

c) (*) Si el 30 % de los requerimientos necesita grabar en disco (y el 70 % restante no), hallar el valor esperado para la cantidad de requerimientos de disco en el lapso de 15 segundos.

24. Para los vectores aleatorios de los ejercicios 1 y 3, hallar

a) el mejor predictor constante de Y y su error cuadrático medio de predicción.

b) el mejor predictor lineal de Y basado en X y su error cuadrático medio de predicción.

PROBABILIDADES Y ESTADÍSTICA (C)

Práctica 5

1. En una línea de producción los productos pasan por 4 procesos sucesivos (preparación, armado, control y embalaje) hasta quedar listos para la venta. Sean X_1, X_2, X_3 y X_4 los tiempos que tarda en cumplirse cada uno de los procesos (en minutos). Se sabe que los cuatro procesos actúan en forma independiente y que $X_1 \sim U(3,5)$, $X_2 \sim N(3,1/4)$, $X_3 \sim \Gamma(6,3)$ y $X_4 \sim U(2,4)$. Sea Y la variable aleatoria que mide el tiempo que tarda un producto en pasar por toda la línea de producción. Hallar $E(Y)$ y $V(Y)$.

2. a) Sea X una v.a. con distribución desconocida tal que $E(X) = 5$ y $V(X) = 0.1$. Usando la desigualdad de Tchebyshev, determinar una cota inferior para la probabilidad de que X esté entre 4.5 y 5.5.

   b) Sea $X_1, \ldots, X_n$ una muestra aleatoria (v.a. independientes e idénticamente distribuídas), con $E(X_1) = 5$ y $V(X_1) = 0.1$, y sea $\bar{X}$ su promedio. Determinar una cota inferior, cuando $n = 10$, para la probabilidad de que $\bar{X}$ esté entre 4.5 y 5.5.

   c) ¿Qué pasa con la cota hallada en (b) cuando el tamaño de la muestra n tiende a infinito?

3. El número de aviones que aterrizan en un aeropuerto sigue un proceso de Poisson con intensidad de 10 aviones cada 20 minutos. Determinar una cota inferior para la probabilidad de que el número de aviones que aterrizan en un período de 1 hora esté entre 20 y 40.

4. Sea p la probabilidad de que una persona elegida al azar apoye la pena de muerte. Para estimar p se encuesta a n personas, se cuenta la cantidad de ellas que apoya la pena de muerte (X) y se define la frecuencia relativa

$$f_n = \frac{X}{n}$$

   Cuanto más cerca esté f_n de p, mejor estimador será.

   a) Hallar una cota superior para $P(|f_n - p| > 0.1)$ que no dependa de p. ¿ Cómo se puede mejorar la estimación?

   b) ¿A cuánta gente debería encuestarse para que $P(|f_n - p| > 0.1) \leq 0.1$?

5. Sea $X_1,\ X_2, \ldots$ una sucesión de variables aleatorias independientes con $E(X_i) = \mu$ y $Var(X_i) = \sigma_i^2$. Sea $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

   a) Probar que si $\sigma_i^2 = \sigma^2 < \infty$ (todas tienen la misma varianza), entonces $\overline{X}_n \underset{n\to\infty}{\longrightarrow} \mu$ en probabilidad.

   b) Probar que si las varianzas son distintas, pero están acotadas, es decir que existe C tal que $\sigma_i^2 \leq C$ para todo i, entonces $\overline{X}_n \underset{n\to\infty}{\longrightarrow} \mu$ en probabilidad.

   c) ¿ Es cierto el resultado del ítem anterior si las varianzas no están acotadas?

6. Sea $X_1,\ X_2,\ldots$ una sucesión de variables aleatorias independientes con distribución $P(\lambda)$. Sea $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

   (a) Calcular $E(X_i^2)$.

   (b) Verificar que $Var(X_i^2)$ es finita.

   (c) Calcular el límite en probabilidad de Y_n.

7. Una empresa láctea produce una cierta variedad de queso en unidades cuyo peso (en kg.) es una v.a. con media 2 y varianza 0.04.

   a) Calcular en forma aproximada la probabilidad de que 60 quesos pesen más de 122 kg.

   b) ¿Cuántas unidades serán necesarias aproximadamente para satisfacer un pedido de 5000 kg. con probabilidad mayor o igual que 0.95?

   c) ¿En qué cambiarían los resultados obtenidos en a) y b) si el peso fuera una variable con distribución $N(2, 0.04)$?

8. Sea X una v.a. con distribución $Bi(100, 0.8)$. Usando el Teorema Central del Límite, calcular en forma aproximada:

$$P\left(75 \leq X \leq 85\right) \qquad P\left(X \geq 80\right) \qquad P\left(0.7 \leq \frac{X}{100} \leq 0.8\right)$$

9. Para rellenar una zona baja del río llegan diariamente 3 camiones A, B y C llenos de piedras. Los pesos en toneladas de las cargas de los camiones son v.a. independientes con esperanzas 1.8, 3.8, 4.1 y varianzas 0.1, 0.18, 0.25 respectivamente.

   a) ¿Cuál es la probabilidad aproximada de que en 360 días se superen las 3500 t?

   b) ¿Cuánto tiempo será necesario para superar las 3500 t con una probabilidad aproximada de 0.90 por lo menos?

10. Sean $X_1,\ X_2,\ldots,X_n$ v.a. independientes e idénticamente distribuidas con función de densidad

$$f(x) = \frac{2}{x^3}\, I_{[1,\infty)}(x).$$

    Calcular aproximadamente

$$P\left(\prod_{i=1}^{100} X_i > e^{55}\right).$$

11. En cierto juego de azar la probabilidad de ganar es 0.3. Para participar en el mismo se paga un peso y, en caso de ganar, se reciben 5 pesos.

    a) ¿Cuál es la probabilidad aproximada de que en 100 juegos independientes un jugador gane más de 90 pesos?

    b) ¿Cuántas veces tendrá que jugar para ganar más de 90 pesos con probabilidad aproximada mayor o igual que 0.80?

12. Sea $U_1, \ldots, U_n$ una muestra aleatoria con distribución $\mathcal{U}[0,1]$ y sea h una función continua. Consideremos la integral $I = \int_0^1 h(x)\, dx$.

   a) Se define $I_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} h(U_i)$. En base a la Ley de los Grandes Números, explicar por qué I_n puede utilizarse para calcular en forma aproximada el valor de la integral I.

   b) Los siguientes datos son una muestra aleatoria de una distribución $\mathcal{U}[0,1]$:

   0.635 0.074 0.220 0.746 0.826 0.148 0.821 0.759 0.080 0.929

   Usarlos para aproximar
   $$\int_0^1 \frac{1}{1+x^2}\, dx.$$
   Comparar el resultado obtenido con el valor exacto de esta integral.

   c) Extender el método hallado en (a) para el cálculo aproximado de
   $$I = \int_a^b h(x)\, dx$$
   siendo a y b números reales tales que $a < b$.

13. Consideremos un método de compresión de bases de datos basado en el sistema de codificación de Huffman y supongamos que las bases consideradas sólo pueden contener las letras A, B, C y D. En el método de compresión considerado, se almacenan las letras de la siguiente forma: el bit 1 representa la A, el par de bits 00 representa la B y los tríos 011 y 010 representan a C y D respectivamente. Las bases de datos almacenarán 500 letras (A, B, C o D). Sea X_i el número de bits necesarios para almacenar la i-ésima letra. Se puede asumir que las X_i son independientes (es decir, que hay independencia entre los caracteres involucrados).

   Después de cierta investigación se sabe que en las bases de datos consideradas el 70 % de las letras es A, el 12 % es B, el 10 % es C y el 8 % es D.

   a) Hallar la esperanza y varianza de la variable X_i, $i = 1, \ldots, 500$. Calcular la esperanza y la varianza del total de bits utilizados al almacenar las 500 letras.

   b) Calcular en forma aproximada la probabilidad de que más de 720 bits sean utilizados en total para almacenar las 500 letras.

   c) Calcular la probabilidad de que se utilice más de 1.46 bits por letra al almacenar 500 letras.

   d) Repetir el ítem anterior para almacenar n letras. Graficar (por ejemplo, en `R`) esta probabilidad como función de n. ¿Qué se observa?

PROBABILIDADES Y ESTADÍSTICA (C)

Práctica 6

1. Una fábrica de alfajores tiene dos sedes: una en Quilmes y la otra en Pilar. Cada sede empaqueta sus alfajores en cajas con 4 unidades. Denotemos con X a la variable aleatoria que indica el número de alfajores defectuosos en una caja y con Y a la variable aleatoria que indica la sede de la que proviene la caja, Quilmes($Y = 0$) o Pilar ($Y = 1$).

   El archivo *alfajores.txt* contiene datos que se obtienen al examinar cajas al azar producidas por la fábrica. Es decir, realizaciones del vector (X, Y).

   Supongamos que se seleccionaron $n = 500$ cajas al azar producidas por la fábrica.

   Estimar

   a) la probabilidad de que una caja provenga de la sede Quilmes.

   b) la probabilidad de que una caja sea producida por la sede Quilmes y tenga 3 alfajores defectuosos.

   c) la función de probabilidad puntual conjunta del vector (X, Y).

   d) la esperanza y la varianza de X.

   e) la probabilidad de que una caja producida por la sede Quilmes tenga 3 alfajores defectuosos.

   f) la probabilidad de que una caja producida por la empresa y elegida al azar tenga 3 alfajores defectuosos (es decir $X = 3$).

   g) la probabilidad de que una caja con 3 alfajores defectuosos haya sido producida por la sede Quilmes.

   h) la probabilidad de que una caja con 3 alfajores defectuosos haya sido producida por la sede Pilar.

2. Se quiere estudiar la distribución de la duración (en horas) de las lámparas producidas por una fábrica. Se observa la duración de 27 lámparas elegidas al azar de la producción de dicha fábrica y se obtienen los siguientes valores (también pueden encontrarlos en el archivo `lamparas.txt`).

26.43	33.58	65.86	29.18	5.92	13.29	13.54	64.78	56.11
23.60	33.39	100.32	28.04	29.63	2.41	3.17	11.99	6.47
23.59	17.96	32.27	2.09	57.43	15.31	42.85	1.68	49.61

   a) Estimar la probabilidad de que una lámpara producida por esta fábrica dure más de 30 horas.

   b) Hallar y graficar la función de distribución empírica para este conjunto de datos.

c) Completar: Estos datos permiten estimar que el 90 % de las lámparas producidas por esta fábrica dura más de .... horas y el 10 % dura menos de .... horas.

3. El archivo **graduados.txt**, contiene los promedios obtenidos en su carrera de grado de 30 inscriptos en el programa de postgrado del Departamento de Ingeniería Industrial e Investigación Operativa de la Universidad de Berkeley, California.

   a) Calcular la media muestral y la mediana muestral.

   b) Calcular el desvío estándar muestral y la distancia intercuartil.

   c) Realizar un histograma con los datos y superponga la curva de una densidad normal con los parámetros que considere pertinentes.

   d) Realice un boxplot con este conjunto de datos. ¿Cuáles son sus características más sobresalientes? ¿Cómo relaciona lo observado en los gráficos con los valores estimados de media y mediana obtenidos en a)? ¿Hay outliers?

   e) ¿Qué distribución podría suponer que tienen estos datos?

   f) Superponer en el histograma la curva de una densidad apropiada con los parámetros que considere pertinentes. Explorar el comando `density` en R.

   g) ¿Qué otro gráfico conoce que le permitiría verificar si su conjetura es razonable?

4. La siguiente tabla contiene valores de población, en cientos de miles, de las 10 ciudades más pobladas de 4 países en el año 1967. Estos datos se encuentran en el archivo **ciudades.txt**.

Argentina	**EEUU**	**Holanda**	**Japón**
29.66	77.81	8.68	110.21
7.61	35.50	7.31	32.14
6.35	24.79	6.02	18.88
4.10	20.02	2.64	16.38
3.80	16.70	1.75	13.37
2.75	9.39	1.72	11.92
2.70	9.38	1.51	10.71
2.69	8.76	1.42	7.80
2.51	7.63	1.31	7.70
2.44	7.50	1.29	7.00

   a) Construir en paralelo, para facilitar la comparación, un boxplot para los datos de cada país e identificar los puntos extremos en cada uno de ellos.

   b) Comparar los centros de cada población, sus dispersiones y su simetría. ¿Cuál diría que es el país más homogéneamente habitado?

5. El archivo **ingresos.txt** contiene el ingreso mensual de un conjunto de 1000 trabajadores registrados de una ciudad, en miles de pesos.

   a) ¿Cuál es el ingreso mínimo percibido por los trabajadores encuestados? Estimar la proporción de los trabajadores de la ciudad que percibe el ingreso mínimo.

   b) Estimar el ingreso mensual que se necesita para pertenecer al 10 % de trabajadores de la ciudad con ingresos más altos.

   c) Calcular la media muestral, la mediana muestral y la media α-podada con $\alpha = 0{,}10$ (10 %).

   d) Calcular el desvío estándar muestral y la distancia intercuartil.

   e) Realizar un histograma y un boxplot. ¿Cuáles son las características más sobresalientes? ¿Hay outliers?

   f) ¿Cree que los datos podrían provenir de una población con distribución normal?

   g) Discutir con un compañero las ventajas y desventajas de cada medida de posición para describir el centro de los datos.

6. Este ejercicio es para familiarizarse con el uso e interpretación de los QQ–plots.

   a) Generar muestras de tamaño 25, 50 y 100 de una variable con distribución normal. Construir QQ–plots para cada una de ellas. Repetir con diferentes semillas.

   b) Repetir a) para una variable con distribución $\Gamma(5, \frac{1}{2})$.

   c) Repetir a) para $Y = \dfrac{Z}{U}$ donde $Z \sim N(0,1)$ y $U \sim \mathcal{U}(0,1)$ independientes.

   d) Repetir a) para una variable con distribución uniforme.

   e) Repetir a) para una variable con distribución exponencial.

   f) ¿Puede distinguir, en base a los QQ–plots, entre una variable con distribución normal del ítem y las siguientes con distribuciones que no son normales?

7. Generar un conjunto de $n = 10$ datos provenientes de una distribución cualquiera. Llamémoslos $x_1, \ldots, x_n$.

   a) Implementar y graficar la función ℓ_2 que, a cada c le asigna $\ell_2(c) = \sum_{i=1}^{n} (x_i - c)^2$. ¿Para qué valor o valores de c se minimiza la función? Experimentar con otros valores de n. Conjeturar una respuesta y luego hacer la demostración para probar su conjetura.

   b) Implementar y graficar la función ℓ_1 que ahora, a cada c le asigna $\ell_1(c) = \sum_{i=1}^{n} |x_i - c|$. ¿Para qué valor o valores de c se minimiza la función? Experimentar con varios valores de n y conjeturar una respuesta.

8. Sean $X_1, \ldots, X_n$ una muestra aleatoria de una población cualquiera.

a) Sean $\bar{X}$ y $\tilde{X}$ la media y la mediana muestral, respectivamente.

   i) Si se suma una constante c a cada uno de los X_i de la muestra, obteniéndose $Y_i = X_i + c$, ¿cómo se relacionan $\bar{X}$ con $\bar{Y}$ y $\tilde{X}$ con $\tilde{Y}$?

   ii) Si cada X_i es multiplicado por una constante c, obteniéndose $Y_i = cX_i$, responder a la pregunta planteada en (i).

b) Sea $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ la varianza muestral correspondiente a la muestra. Demostrar que:

   i) Si $Y_i = X_i + c$, con c constante, entonces $S_Y^2 = S_X^2$.

   ii) Si $Y_i = cX_i$, con c constante, entonces $S_Y^2 = c^2 S_X^2$.

   iii) $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\bar{X}^2$.

PROBABILIDADES Y ESTADÍSTICA (C)

PRÁCTICA 8

Sobre la notación: A lo largo de esta materia, utilizamos, indistintamente, $\overline{x}_n$, $\overline{x}$, $\overline{X}_{\text{obs}}$ para denotar promedios. Un abuso de notación análogo usamos para los valores observados de otros estimadores.

1. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal.

   (a) Encontrar un intervalo de confianza de nivel exacto $1 - \alpha$ para la media cuando la varianza es conocida.

   (b) Con cierto instrumento se realizan 20 mediciones de una magnitud física μ. Cada observación es de la forma $X_i = \mu + \epsilon_i$, donde ϵ_i denota el error aleatorio de la i-ésima medición. El promedio de los valores obtenidos es $\overline{x} = 25.01$.

   Supongamos que los errores de medición tienen distribución normal con media cero y varianza 0.36.

   i) Hallar el intervalo de confianza estimado de nivel exacto 0.95 para el valor de la magnitud física de interés.

   ii) Si se quisiera que la longitud del intervalo hallado en i) fuera a lo sumo 0.05, ¿cuántas mediciones deberían realizarse?

   iii) Suponga ahora que la varianza del error es 0.6. ¿cuántas mediciones deberían realizarse para construir un intervalo de confianza exacto de nivel 0.95 para μ cuya longitud fuera a lo sumo 0.05?

   iv) Compare los resultados obtenidos en los items ii) y iii) e interprete la diferencia.

   (c) En `https://probac2019.shinyapps.io/datos_practica8_download/` encontrará una aplicación para generar resultados simulados de este experimento. Genere un conjunto de datos de tamaño 10 y rehaga el item b) i). Luego genere un conjunto de datos del tamaño hallado en el item b) ii), calcule el intervalo de confianza estimado y verifique que su longitud es menor que 0.5.

2. Consideremos variables aleatorias $X_1, \ldots, X_n$ con distribución $N(\mu, \sigma^2)$. En este ejercicio estudiaremos la distribución de $T_n = (\overline{X}_n - \mu)\sqrt{n}/s$ para diferentes valores de n. Para ello, para diferentes valores de n, y un μ y un σ dados, simularemos 1000 conjuntos de n datos con distribución $N(\mu, \sigma^2)$ y luego calcularemos el estadístico T_n para cada conjunto de datos. Obtendremos así 1000 realizaciones de la variable $T_n = (\overline{X}_n - \mu)\sqrt{n}/s$. Finalmente, haremos boxplots, qqplots, e histogramas de las 1000 realizaciones de T_n para estudiar su densidad.

   Generar dos valores con distribución $N(1, 4)$ y computar el valor que toma el estadístico T_n para esos valores observados, $T_{n\,obs}$. Replicar 1000 veces, guardando los resultados en un vector de longitud 1000.

(a) Hacer un boxplot y un qqplot del conjunto de valores de $T_{n\,obs}$. ¿Qué se observa ?

(b) A partir de los valores replicados de $T_{n\,obs}$ realizar un histograma. ¿Qué características tiene este histograma?

(c) Al histograma del item anterior, superponerle el gráfico de la densidad normal estándar. Superponer también el gráfico de la verdadera densidad de la variable T_n. ¿Qué observa?

(d) Con el comando `density`, dibujar la densidad estimada de T_n. En el mismo gráfico, superponer la verdadera densidad de T_n y la curva normal estándar. ¿Qué observa?

Repetir para $n = 4, 20, 30, 500$ y comparar los resultados.

3. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal.

   (a) Encontrar un intervalo de confianza de nivel exacto $1 - \alpha$ para la media cuando la varianza es desconocida.

   (b) Repetir la parte i) del item b) del Ejercicio 1, suponiendo que la varianza es desconocida y el desvío estándar muestral es $s = 0.76$.

4. El rendimiento anual de almendros (en kg.) en parcelas cultivadas en cierta región es una variable aleatoria con distribución normal de media μ y varianza σ^2. Se observa el rendimiento anual de $n = 5$ parcelas elegidas al azar y se obtiene una media muestral de 525 kilogramos y un desvío muestral de 10 kilogramos.

   (a) En base a los datos obtenidos, obtenga una estimación por intervalos para μ utilizando un procedimiento de nivel 0.95.

   (b) Decidir si son verdaderas o falsas las siguientes afirmaciones:

      i. Aproximadamente el 95% de las parcelas de esta región tiene un rendimiento que pertenece al intervalo hallado en a), en base a la muestra observada.

      ii. La probabilidad de que el intervalo hallado en a) contenga a μ es 0.95.

      iii. Si se extraen muchas muestras de tamaño 5 de manera independiente y para cada una se calcula el intervalo de confianza como en a), aproximadamente el 95% de estos intervalos contendrá a μ.

5. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal.

   (a) Encontrar un intervalo de confianza de nivel exacto $1-\alpha$ para la varianza cuando la media es conocida. SUGERENCIA:: probar que $\sum_{i=1}^{n}(X_i - \mu)^2/\sigma^2 \sim \chi^2_n$.

   (b) En un aserradero se cortan varillas de madera cuya longitud es una v.a. con distribución normal. Se miden 25 varillas elegidas al azar, obteniéndose las siguientes longitudes

   Hallar el intervalo de confianza estimado de nivel exacto 0.90 para la varianza verdadera, suponiendo que $\mu = 185$.

176.50	191.50	186.90	181.10	195.70
188.10	187.40	185.10	176.90	191.20
193.80	187.00	179.00	173.00	184.40
199.60	190.40	206.80	193.00	177.10
201.00	192.50	176.60	180.10	186.40

6. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal.

   (a) Encontrar un intervalo de confianza de nivel exacto $1-\alpha$ para la varianza cuando la media es desconocida.

   (b) Repetir el item b) del Ejercicio 5 suponiendo que μ es desconocida.

7. (a) Definir una función que, tenga por input un conjunto de datos $x = (x_1, \ldots, x_n)$ provenientes de una distribución normal y un nivel de confianza $1-\alpha$ y devuelva un intervalo de confianza de nivel $1-\alpha$ para la media de la normal de la que provienen los datos. ¿Cómo cambiaria la función si ahora queremos aplicarla a normales con varianza conocida? ¿Y si ahora queremos aplicarla a muestras grandes pero con distribución desconocida?

   (b) Elegir un valor de μ y un valor de σ. Generar 1000 conjuntos de datos de n valores cada uno con distribución $N(\mu, \sigma^2)$ y calcular intervalos de confianza para μ asumiendo que μ es desconocido y σ es conocido. Repetir para diferentes valores de n entre 3 y 1000. ¿Qué proporción de intervalos contiene al verdadero valor de μ?

   (c) Rehacer el item anterior calculando también, para cada conjunto de datos, el intervalo de confianza para μ con σ desconocido. Para valores grandes de n, ¿de qué otra forma podría calcula los intervalos de confianza? Calcular los tres intervalos para cada muestra y compararlos.

   (d) Rehacer los dos items anteriores pero, en lugar de datos provenientes de una normal, generar datos provenientes de una $U(0,1)$. Interpretar.

8. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población $E(\lambda)$.

   (a) Probar que $2\lambda\sum_{i=1}^{n} X_i$ tiene distribución χ^2_{2n}.

   (b) Hallar un intervalo de confianza para λ de nivel exacto $1-\alpha$.

   (c) ¿Cuál sería el intervalo de confianza de nivel exacto $1-\alpha$ para $E(X_1)$? ¿Cuál es su longitud esperada?

   (d) Aplicar b) a los datos del Ejercicio 4 d) de la Práctica 7, con nivel $1-\alpha = 0.95$.

   (e) Hallar un intervalo de nivel asintótico $1-\alpha$ para λ.

9. Se desea conocer la opinión de los ciudadanos de cierta población acerca de una propuesta política. Para ello, se realiza una encuesta con tres posibles respuestas: a favor, en contra o indeciso.

   (a) Hallar un intervalo de confianza de nivel asintótico 0.90 para la proporción de votantes que se oponen a la propuesta. ¿Qué supuestos debe hacer?

   (b) Se realiza la encuesta a 1000 ciudadanos, obteniéndose como resultado que 200 están a favor 600 en contra y 200 están indecisos. Hallar el intervalo de confianza estimado de nivel asintótico 0.90 para la proporción de votantes que se oponen a la propuesta. Interpretar el resultado.

   (c) ¿Cuántos votantes deberían encuestarse para que la longitud del intervalo obtenido fuese menor o igual que 0.02?

   (d) En `https://probac2019.shinyapps.io/datos_practica8_download/` encontrará una aplicación para generar resultados simulados de este experimento. Genere un conjunto de datos de tamaño 1000 y rehaga el item (b). Luego genere un conjunto de datos del tamaño hallado en el item (c), calcule el intervalo de confianza estimado y verifique que su longitud es menor que 0.02.

10. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población $Bi\,(k, \theta)$.

    (a) Hallar un intervalo de confianza de nivel asintótico $1-\alpha$ para θ, siendo k conocido.

    (b) Encontrar una cota superior para la longitud del intervalo hallado en a).

11. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población $P\,(\lambda)$.

    (a) Hallar un intervalo de confianza de nivel asintótico $1 - \alpha$ para λ.

    (b) Aplicar a) a los datos del Ejercicio 5 d) de la Práctica 7, con $\alpha = 0.05$.

12. (a) Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución $U[0, \theta]$. Hallar un intervalo de confianza de nivel exacto $1 - \alpha$ para θ.
    Sugerencia: Encontrar la distribución de $\max\limits_{1 \leq i \leq n} (X_i)\,/\theta$.

    (b) Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con densidad

    $$f(x; \theta) = e^{-(x-\theta)} I_{[\theta,\infty)}(x)$$

    Hallar un intervalo de confianza de nivel exacto $1 - \alpha$ para θ.
    Sugerencia: Encontrar la distribución de $\min\limits_{1 \leq i \leq n} (X_i) - \theta$.

    (c) (**Opcional**) En los casos a) y b) obtener el intervalo de menor longitud esperada.

13. Sean X e Y dos variables aleatorias con distribución normal con la misma varianza: $X \sim N(\mu_1, \sigma^2)$ e $Y \sim N(\mu_2, \sigma^2)$. Considere las siguientes observaciones para X:

$$0.44,\ -1.63,\ 2.59,\ 1.54,\ 0.45,\ -0.13,\ -2.76,\ -1.53$$

y para Y:

$$0.06,\ -0.24,\ 4.65,\ 2.27,\ 3.88,\ 2.35,\ 3.92,\ -0.73.$$

(a) Calcule el intervalo de confianza estimado de nivel 0.95 para μ_1.

(b) Calcule el intervalo de confianza estimado de nivel 0.95 para μ_2.

(c) Calcule el intervalo de confianza estimado de nivel 0.95 para $\mu_1 - \mu_2$.

PROBABILIDADES Y ESTADÍSTICA (C)

Práctica 9

1. Un productor agropecuario cultiva una variedad de soja cuyo rendimiento por hectárea (ha.) puede considerarse una variable aleatoria con distribución $N(37, 25)$. Un vendedor de granos pone a la venta una nueva variedad de soja, diciendo que tiene un rendimiento mayor. El productor agropecuario realizó una compra de esta nueva variedad y quiere asegurarse de que tienen un rendimiento mayor que las semillas que compra habitualmente. Con ese fin decide cultivar 10 parcelas de 1 ha., observar sus rendimientos y aplicar un test de hipótesis de nivel 0.05. Suponiendo que el rendimiento por ha. de la nueva variedad de soja tiene una distribución normal con varianza 25, resolver los siguientes ítems.

   (a) Definir las variables aleatorias y los parámetros involucrados en este problema y plantear el test de hipótesis que debe realizar el productor agropecuario.

   (b) El productor cultiva las 10 parcelas y obtiene los siguientes rendimientos:
   37 - 39.50 - 41.70 - 42 - 40 - 41.25 - 43 - 44.05 - 38 - 38.50
   ¿Cuál es la conclusión del test en base a estos datos observados? Calcular la probabilidad del error de tipo II suponiendo que el verdadero rendimiento medio de la nueva variedad de soja es de 40 toneladas por ha.

   (c) Encontrar el test para el caso en que se cultiven n parcelas (con igual nivel).

   (d) Determinar el número n de parcelas a cultivar para que la probabilidad del error de tipo II sea menor o igual que 0.05 suponiendo que el verdadero rendimiento medio de la nueva variedad de soja es de 40 toneladas por ha.

   (e) Calcular la función de potencia $\pi(\mu)$ del test planteado en c), verificar que es creciente para $\mu \in \mathcal{R}$ y deducir que este test sirve también para testear

   $$H_0 : \mu \leq 37 \qquad H_1 : \mu > 37.$$

2. En la construcción de un edificio debe usarse un concreto que tenga una tensión media mayor a 300 psi. Para verificar si el concreto preparado a partir del cemento "Loma Blanca" cumple con este requerimiento, se realizan 15 mediciones en forma independiente de la tensión de este concreto. Se observa una media muestral de 304 psi y un desvío estándar muestral de 10 psi. El constructor está dispuesto a correr un riesgo del 5% de comprar cemento "Loma Blanca" cuando éste produce un concreto que no cumple con las especificaciones. Suponiendo que los datos están normalmente distribuidos:

   (a) Plantear el test correspondiente. ¿Qué decisión se toma?

   (b) Acotar el p.-valor.

3. Se diseñó un nuevo sistema de riego de manera tal que el desvío del tiempo de activación sea menor que 6 segundos. Se lo prueba 11 veces, obteniéndose los siguientes tiempos de activación:

   27 41 22 27 23 35 30 24 27 28 22

   Suponiendo que el tiempo de activación (en segundos) es una v.a. con distribución normal:

   (a) ¿Usted decidiría, a nivel 0.05, que el sistema cumple la especificación?

   (b) Acotar el p.-valor

   (c) Encontrar la función de potencia $\pi(\sigma^2)$, implementarla en R y graficarla. ¿Es creciente o decreciente? ¿Cuánto vale en $\sigma = 6$?

4. Se supone que 1 de cada 10 fumadores prefiere la marca A. Después de una campaña publicitaria en cierta región de ventas, se entrevistó a 200 fumadores para determinar la efectividad de la campaña. El resultado de esta encuesta mostró que 26 personas preferían la marca A.

   (a) ¿Indican estos datos, a nivel aproximado 0.05, un aumento en la preferencia por la marca A?

   (b) Calcular el p-valor aproximado.

   (c) ¿Cuál es la probabilidad aproximada de decidir que la campaña publicitaria no fue efectiva, cuando en realidad la proporción de preferencia por la marca A después de la campaña es 0.15?

   (d) Implementar la función potencia y graficarla. ¿Qué características tiene esta función? Interpretar.

   (e) ¿Qué tamaño de muestra debería tomarse para que la probabilidad de c) fuese a lo sumo 0.05?

   (f) Decidir si las siguientes afirmaciones son V o F . Justificar.

      (i) Si la verdadera proporción de fumadores que prefieren la marca A es mayor que 0.15, la probabilidad de que este test rechace H_0 es mayor que 0.725.

      (ii) Cuánto más pequeña es la verdadera proporción de fumadores que prefieren la marca A más improbable es que el test rechace H_0.

      (iii) $\pi(0.1) = 0.05$.

      (iv) La probabilidad de error de tipo II es una función creciente de p, donde p es la verdadera proporción de fumadores que prefieren la marca A.

5. Se supone que el tiempo de duración de cierto tipo de lamparitas tiene distribución exponencial. Una fábrica garantiza que el tiempo medio de vida de las lamparitas que produce es mayor que 50 días, y la empresa vendedora quiere asegurarse que la producción satisface las especificaciones antes de sacarla a la venta. Para ello toma

al azar una muestra de 40 lamparitas y observa el tiempo de duración de las mismas, obteniendo un promedio de 53 días.

(a) Proponer un test de nivel exacto con hipótesis nula simple para este problema de manera tal que la probabilidad de que la empresa decida no vender si no se satisfacen los requerimientos sea del 0.95.

(b) ¿Qué decisión se toma en base a los tiempos observados?

(c) Repetir a) usando un nivel aproximado.

(d) Calcular la función de potencia aproximada para el test del ítem anterior. ¿Este test conserva el nivel aproximado 5% si ampliamos la hipótesis nula, reemplazándola por una hipótesis compuesta?

(e) Utilizando el test aproximado, ¿qué probabilidad tiene la empresa de no sacar la producción a la venta, si el promedio de vida verdadero es 52 días?

(f) ¿Cuál debería ser el tamaño de la muestra para que la probabilidad de d) fuese 0.1?

6. Se desea determinar si un dado de seis caras está cargado o no. Para ello se arroja el dado 1000 veces obteniéndose los resultados contenidos en el archivo **dado.txt**, que se encuentra en la página web de la materia.

   Estamos interesados en testear las hipótesis H_0 : El dado es equilibrado vs. H_1 : El dado está cargado. Para ello consideramos distintos tests:

   (a) Sea X = cantidad de veces que el resultado del dado es par. Plantear las hipótesis y determinar la región de rechazo para un test de nivel aproximado $\alpha = 0.05$ basado en X. Para el conjunto de datos obtenido, ¿cuál es el valor p aproximado?, ¿cuál es la decisión?.

   (b) Sea Y = cantidad de veces que el resultado del dado es menor o igual que 3. Plantear claramente las hipótesis y determinar la región de rechazo para un test de nivel aproximado $\alpha = 0.05$ basado en Y. Para el conjunto de datos obtenido, ¿cuál es el p-valor aproximado?, ¿cuál es la decisión?.

   (c) Sea W = cantidad de veces que el resultado del dado es exactamente 3. Plantear claramente las hipótesis y determinar la región de rechazo para un test de nivel aproximado $\alpha = 0.05$ basado en W. Para el conjunto de datos obtenido, ¿cuál es el p-valor aproximado?, ¿cuál es la decisión?.

   (d) Sea U = cantidad de veces que el resultado del dado es exactamente 2. Plantear claramente las hipótesis y determinar la región de rechazo para un test de nivel aproximado $\alpha = 0.05$ basado en U. Para el conjunto de datos obtenido, ¿cuál es el p-valor aproximado?, ¿cuál es la decisión?.

   (e) En base a los resultados obtenidos. ¿Qué decisión tomaría? ¿Está cargado el dado?

7. En cada caso indique si la afirmación es verdadera o falsa y justifique:

   (a) El nivel de significación de un test es igual a la probabilidad de que la hipótesis nula sea cierta.

   (b) Un Error de tipo II es más grave que un Error de tipo I.

   (c) Si el p-valor es 0.3, el test correspondiente rechazará al nivel 0.01.

   (d) Si un test rechaza al nivel de significación 0.06, entonces el p-valor es menor o igual a 0.06.

   (e) Si un intervalo de confianza de nivel 0.99 para la media μ de una distribución normal calculado a partir de una muestra da como resultado (2.1, 3.3), entonces el test para las hipótesis $H_0 : \mu = 3$ vs $H_1 : \mu \neq 3$ basado en los mismos datos rechazará la hipótesis nula al nivel 0.01.